Supplemental Figures

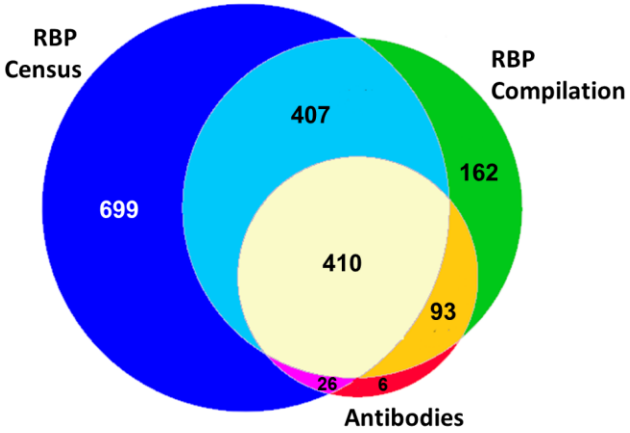

Figure S1. Number of overlapping RBPs between RBP lists and antibodies, related to Figure 1.
Venn diagram showing the number of RBPs overlapping between the RBP Compilation (1072 RBPs), RBP Census from Gerstberger et al (1542 RBPs) and the list of 535 unique RBPs for which 700 antibodies were evaluated for IP in K562. It is noted that our RBP Compilation focused on pre-mRNA binding proteins.

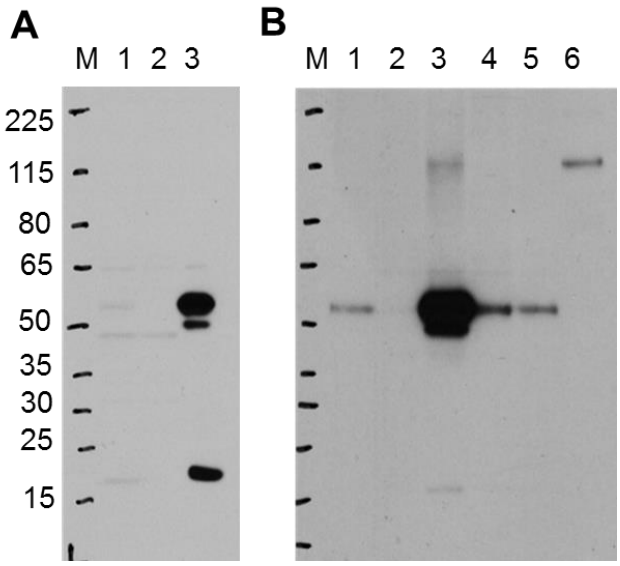

Figure S2. Antibody IP-scores are cell-type dependent, related to Figure 1.
Both panels A and B are IP validations of A300-864A (Bethyl Laboratories) antibody against RBFOX2 protein. (A) IP validation using K562 whole cell lysate. The validation was scored as '1IP' as the protein band could not be detected in the Input lane. (B) IP validation using HepG2 whole cell lysate and the validation was scored as '1'. In both (A) and (B), lane 1 is 2.5% whole cell lysate, lane 2 is 2.5% of supernatant after IP and lane 3 is 50% of IP pull down using RBFOX2 antibody. Lanes 4-6 are the same but using Rabbit normal IgG.

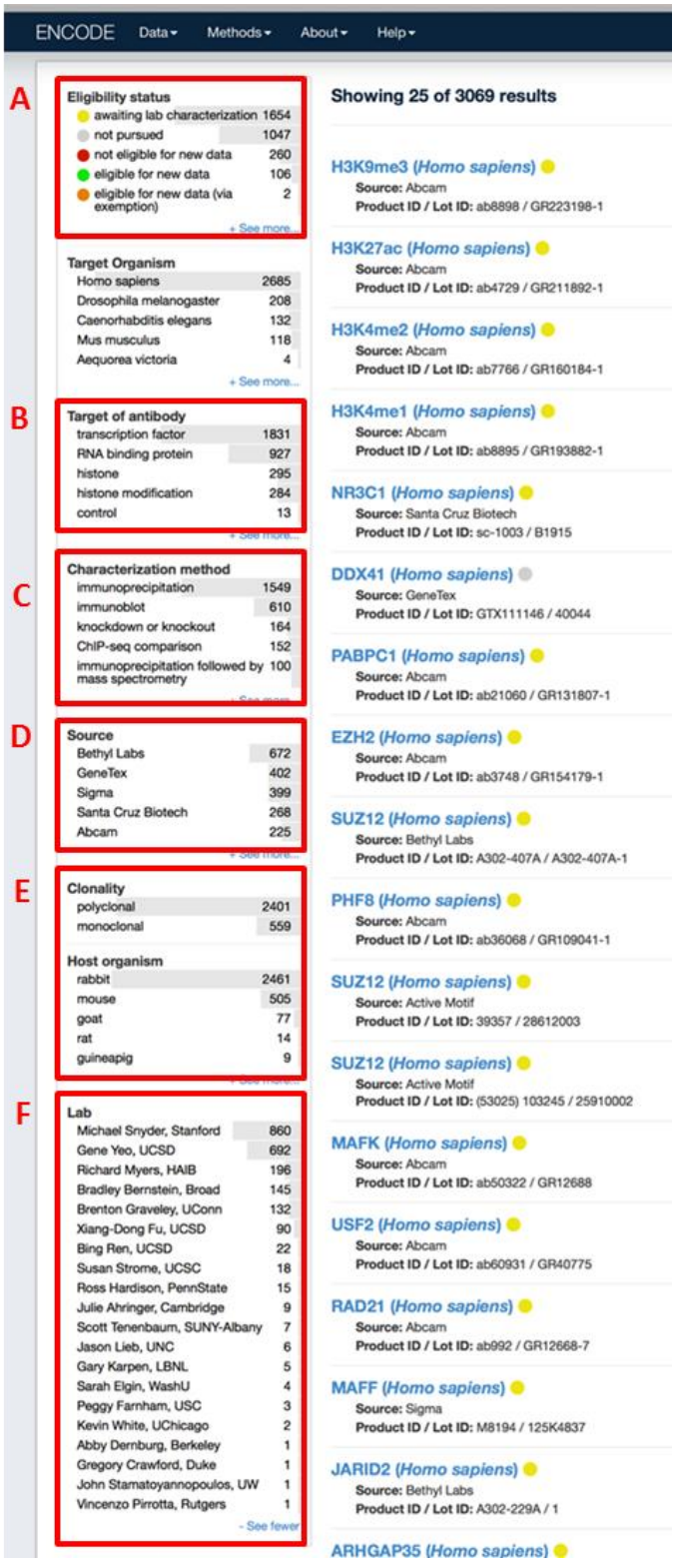

Figure S3: Browsing antibody pages in the ENCODE portal, related to Figure 2.

Users can browse the Data → Antibodies page in the ENCODE portal for choosing antibodies. The results of antibodies can be a shortlisted based in the filtering criteria on the left side panels as described below.

A. Eligibility Status:

This criterion chooses antibodies based on how the validation results are classified according to the ENCODE standards as described in main text. For further details about how the antibodies are classified based on the validation results, refer https://www.encodeproject.org/help/antibody_characterization_process/

B. Target of antibody:

This criterion filters antibodies based on whether they target a RBP or TF etc.

C. Characterization method:

The results can be sorted by the validation methods like immunoblot (WB) or immunoprecipitation (IP) etc.

D. Source:

Antibodies can also be sorted by the vendors like Bethyl lab, Abcam etc.

E. Clonality and Host organism:

Antibody pages can also be sorted by clonality and host species like mouse monoclonal vs rabbit polyclonal etc.

F. Lab:

And finally the results can also be sorted based on which lab characterized the antibodies.

The results can be sorted based on multiple criteria by choosing more than one option in the left side panel. For example, to view all the Bethyl antibodies characterized by Yeo lab, the user has to click ‘Gene Yeo, UCSD’ option in panel F and then click ‘Bethyl Labs’ in panel D.

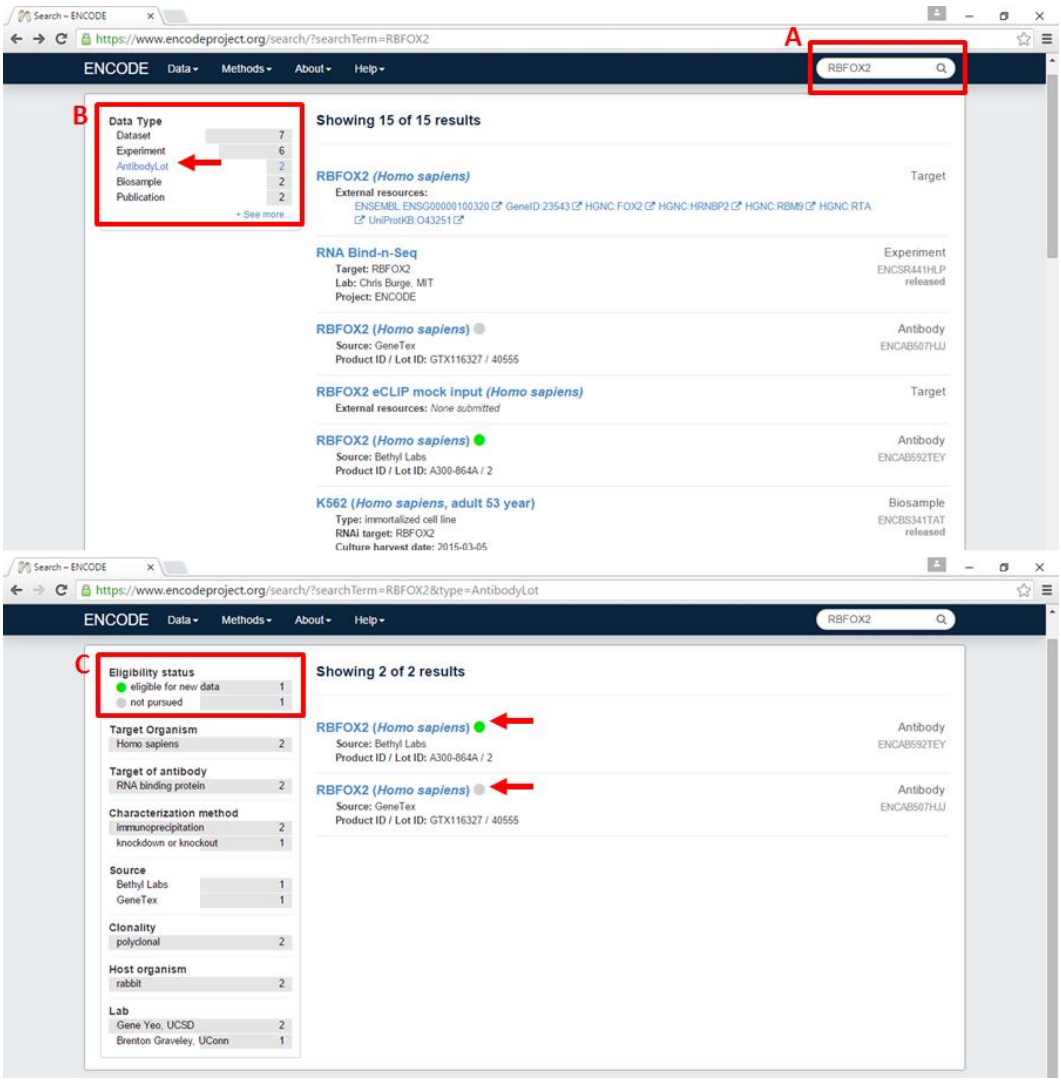

Figure S4: Searching for antibodies using text search in the ENCODE portal, related to Figure 2.

Users can search for antibodies by the target protein name or by the catalogue number using the 'text-search box' as shown in Panel A ('RBFOX2' for example). ENCODE portal will show all the pages like 'Experiment' (showing all the ENCODE related experiment linked with search term), 'AntibodyLot', 'Biosample' etc. Clicking the 'AntibodyLot' as shown by the red arrow in Panel B will bring the entire list of antibody pages associate with the search term. If more than one antibody is characterized for the protein, then the results will be sorted based on antibodies eligibility criteria like 'eligible for new data', 'not pursued', 'awaiting lab characterization' etc with an accompanying color circle as in Panel C. Green circle indicates a well characterized antibody that is most preferred for a CLIP or RIP experiment, Yellow circle is for antibodies that are waiting further validations, Grey circle for partially characterized antibodies and Red circle indicate antibodies that failed validations and must not be used for further experiments.

Supplemental Tables

Table S1. RBP Compilation, related to Figure 1.

Sheet 1 is the list of 476 proteins that are compiled based on presence of the RNA Binding Domains (RBDs).

Sheet 2 is the '1072 RBP Compilation' list which is the union of 476 RBPs identified by RBDs and the 845 proteins identified in the RBP interactome capture study (Castello et al., 2012).

Sheet 3 is the list of ENSEMBL IDs for the 1072 RBP Compilation and associate Pfam domain IDs resulted from Biomart tool (http://www.ensembl.org/biomart/) search.

Table S2. RBP Antibody Details, related to Figure 1.

Sheet 1, 2 and 3 contain metadata like host species, clonality, antigen information, purification method etc for antibodies from Bethyl, MBLI and GeneTex companies respectively. Each sheet may vary in number of columns and specific information based on disclosure policy of the vendors.

Table S3. RBP shRNA Details, related to Figure 3.

Information about all RBP shRNAs acquired including the target genes.

Table S4. Results of the IP-WB validations, related to Figure 1.

Sheet 1 is the list of all 700 IP-WB experiments performed using K562 cells, containing catalog and lot numbers of antibodies, Uniprot ID and expected MW of RBPs, IP score, ENCAB ID and antibody status for each antibody.

Sheet 2 is distribution for 700 antibodies by IP score as well as by vendor wise breakdown.

Sheet 3 is the Uniprot IDs for 535 unique RBPs for which at least one antibody was tested in K562.

Sheet 4 is the list of 438 IP-grade antibodies.

Sheet 5 is the list of ENSEMBL, HGNC symbol and Pfam IDs for 365 unique RBPs that have IP-grade antibodies.

Sheet 6 is the list of 332 Pfam IDs associated with 365 RBPs.

Sheet 7 is the total occurrence of each Pfam domain type among 365 RBPs.

Table S5. Summary of the shRNA-KD validation experiments for 370 shRNA constructs, related to Figures 3 and 4. Information presented includes the TRC number of shRNA plasmids, sequences of primers used for RT-qPCR validation and catalog numbers of antibodies used in the KD-WB characterization.

Table S6. Details of the immune fluorescence imaging experiments, related to Figure 5.

Information including primary antibody concentration, exposure time and holistic classification of IF image quality for antibodies are presented.

Supplemental Experimental Procedures

Immunoprecipitation followed by Western blot (IP-WB)

K562 cells (ATCC CCL-243, lot 59300853) grown in RPMI 1640 (Life Technologies, 11875119) with 10% FBS (Life Technologies, 26140079) and 1% Pen-Sterp (Life Technologies, 15140163) were flash frozen into $5x10^5$ aliquotes and stored at -80˚C. Cell pellets were thawed on ice and resuspended in 450µl Lysis buffer containing 50 mM Tris-HCl, pH 7.4; 100 mM NaCl; 1% NP-40 (IGEPAL CA-50); 0.1% SDS; 0.5% sodium deoxycholate with protease inhibitor (Roche, 11836170001). Cell lysate left in ice for at least 15 minutes for complete cell lysis and then sonicated using Bioruptor for 5min with 30sec on/off pulses in cold water bath. Cell lysate spun at 18000g for 20min in a refrigerated centrifuge and the clear supernatant was stored in new tube on ice for IP. 0.625mg per tube of Dynabeads (anti-mouse (11202D) or anti-rabbit IgG (11204D) secondary or Protein A (10002D) coated beads) washed twice with 900µl cold Lysis buffer and resuspend beads in 250µl of Lysis buffer. 50µl of the washed beads were used to preclear the whole cell lysate by rotating for 30min at RT. With remaining 200µl of beads, 5ug of antibody specific to RBP of interest was added and rotated for at least 1hr at RT for antibody-bead coupling. After preclearing, tubes were kept on magnetic stand (Dynamag, 12321D) and clear lysate transferred to new tube. 25µl (5%) of this precleared lysate was stored at -20˚C as *Input sample* for WB later. After one hour of antibody-bead coupling, excess antibodies were removed and beads were washed three times with 900µl cold lysis buffer on magnetic stand. Precleared cell lysate was then added to the antibody coupled beads and rotated at 4˚C. After overnight immunoprecipitation at 4˚C, tubes were kept on magnetic rack, 25µl (5%) of clear supernatant/ flow through was saved as *IP-sup sample* for WB later. Excess flow through was discarded and beads were washed two times in 900µl cold **High Salt Wash** buffer containing 50 mM Tris-HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate. Beads were then washed two more times with 900µl cold **PNK Wash buffer** containing 20 mM Tris-HCl pH 7.4, 10 mM MgCl2, 0.2% Tween-20. Washed beads were resuspended in 26µl PNK Wash buffer, 10µl 4X LDS sample buffer (Novex, B0008) and 4µl 1M DTT. With the 25µl Input and IP-Sup samples saved previously, 10µl 4X LDS sample buffer and 4µl 1M DTT were added and mixed. All the samples were denatured by boiling at 70˚C for 10min with shaking at 1200 rpm (in Thermomixer). After boiling, samples were immediately placed on ice cold magnetic stand and liquid was removed for gel loading. Samples were loaded on appropriate denaturing 15 well, 1.5mm thick, 4-12% Bis-Tris gel (Novex, NP0336BOX) for RBPs with MW less than 225kDa and 8% Tris-Glycine (Novex, EC6018BOX) for RBPs with MW more than 225kDa. Spectra Broad (Thermo, 26634) or High range (Thermo, 26625) MW markers were used depending on MW of RBP. Samples were run at 150V for 1.30hr using MOPS-SDS running buffer (NP0001). Protein bands were then transferred on to PVDF membrane (BioRad, 1620177) using NuPAGE transfer buffer (NP00061) with 10% methanol at 200mA constant current for 2hr in cold room. Transfer efficiency was checked by Ponceau staining and blots were blocked with 5% skim milk (Genesee Scientific, 20-241). Membranes were then incubated overnight at 4˚C with 0.2-0.5ug/ml (usually 1:2000-1:5000) of the same antibody used for IP as primary antibody. After primary antibody

incubation, membranes were washed three times in 1X TBST for 10min and incubated for 1-3hr at RT with appropriate secondary TrueBlot HRP antibody (Rockland, 18-8816-33, 18-8815-33). After secondary incubation, membranes were washed three times in TBST and developed with ECL (Pierce, 32106) or ECL Plus (Pierce, 32132) substrates depending on the abundance of RBP. Protein marker bands and edges of the membrane were marked down in the photo-developer film which was scanned to generated images. IP results were scored based on the following scoring schema.

1: Only one band, that deviates less than 20% from the expected molecular weight and detected in input lane. The same band is also enriched upon immunoprecipitation with band intensity greater than or equal to that in the Input lane.
1IP: One band is detected only upon immunoprecipitation enrichment that deviates less than 20% from the expected molecular weight. The same band is NOT detected in the input lane due to expression and/or detection level thresholds.
1MW: Only one band is detected in the input and immunoprecipitation lanes, but the observed molecular weight deviates more than 20% from the expected molecular weight. The same band is also enriched upon immunoprecipitation with a band intensity greater than or equal to that in the input lane.
1MB: One prominent band is detected in input and immunoprecipitation lanes that deviates less than 20% from the expected molecular weight. This band is also enriched upon immunoprecipitation with band intensity greater than or equal to that in the input lane. In addition, there are multiple bands *below* the expected molecular weight, which are also detected in the input lane, and/or enriched upon immunoprecipitation.
0.5MB: One prominent band is detected in input and immunoprecipitation lanes that deviates less than 20% from the expected molecular weight. The band is enriched upon immunoprecipitation with an intensity greater than or equal to that in the Input lane. In addition, there are multiple bands *above* the expected molecular weight that are also detected in the input lane and/or enriched upon immunoprecipitation.
0.5: Only one band is detected in input lane and deviates less than 20% from the expected MW. The same band is *poorly* enriched up on immunoprecipitation with band intensity *less than* the input lane.
0WB: The antibody does not enrich the protein upon immunoprecipitation but detects band at the expected MW.
0: The antibody neither detects bands in input lane nor enriches upon IP using K562 lysate.

Analysis of protein domains

The ENSEMBL Biomart webportal (http://www.ensembl.org/biomart/) was used to search for ENSEMBL and Pfam IDs using Uniprot IDs as query. If a RBP has multiple copies of the same Pfam domain then that ID was counted only once for that RBP and if a RBP has more than one type of Pfam domain all of them were counted.

Production of shRNA lentiviral particles

0.8-1x106 293 T cells (catalog number: CRL-11268, ATCC) were plated in each well of 6-well plate with 10 % FBS (catalog number: 30-2020, ATCC), DMEM (catalog number: 11995-065, Life technologies) medium without penicillin and streptomycin. The cells were incubated overnight at 37˚C and were typically 70-80% confluent. A cocktail for each transfection was assembled in a polypropylene tube containing 500 ng pLKO-shRNA, 500 ng psPAX2 Packaging DNA, 50 ng PMD2.G Envelope DNA, and serum-free OPTI-MEM to 100 µl. 3.1 µl of FuGENE HD Transfection reagent (Catalog number: E2311, Promega) was added to the tube (FuGENE:DNA=3:1), incubated for 20 minutes at room temperature, the DNA mix was gently added dropwise to the cells, followed by incubation at 37˚C for 12-15 hr. The next day, the media was changed to remove the transfection reagent, and the cells were washed with PBS once and 1.5 ml fresh media +10% FBS + penicillin/streptomycin was added. After incubation at 37˚C for four days, the media was harvested from cells, stored at 4˚C, and 1.5 ml of fresh media was added and incubated overnight at 37˚C. The media was harvested and pooled with the media collected on day 4, spun at 1250 rpm for 5 min to remove cells and the viral stock stored at -40˚C.

qRT-PCR Lentivirus Titration Assay

Lentiviral titrations were performed using the qPCR Lentivirus Titration kit from Applied Biological Materials Inc. (Catalog Number LV900). 2 µl of the viral supernatant was added to 18 µl of Virus Lysis buffer and incubated at room temperature for 3 minutes. qRT-PCR reactions were assembled to include 12.5 µl of 2X qRT-PCR Mastermix, 2.5 µl of the viral lysate, and 10 µl of the reagent mix. qRT-PCR was performed by incubation at 42˚C for 20 minutes, 95˚C for 10 minutes, followed by 40 cycles of 95˚C for 15 seconds and 60˚C for 1 minute. Viral titers were calculated from the Ct values by using the Applied Biological Materials Inc. on-line lentiviral titer calculator at http://www.abmgood.com/High-Titer-Lentivirus-Calculation.html

Lentiviral Transduction of K562 and HepG2 cells

K562 cells (ATCC CCL-243 (lot 59300853)) were grown in 500 ml RPMI-1640 with glutamine medium (Hyclone, SH30027.01), 50 ml Fetal Bovine Serum (FBS) (10% Final Concentration) (Hyclone, SH30071.03), and 5 ml Pen-Strep (1% Final Concentration) (Invitrogen, 15140-163). HepG2 cells (ATCC HB8065 (lot 59635738)) were grown in 500 ml DMEM (HyClone, SH30022.01), 50 ml Fetal Bovine Serum (FBS) (10% Final Concentration) (Hyclone, SH30071.03), and 5 ml Pen-Strep (1% Final Concentration) (Life Technologies, 15140122). Frozen stocks of cells were thawed by gentle agitation in a 37˚C water bath and the cells into the growth medium and centrifuged at 1000 rpm for 5 minutes. The cell pellets were resuspended in fresh growth medium and incubated at 37˚C in a 5% CO2 air atmosphere incubator. The growth medium was changed every 2 to 3 days and the cells split when the cell density reached 70−80% confluence. For transductions, $5x10^5$ cells were plated into each well of 6-well plates and incubated for overnight such that the cells were 50−60% confluent. The media was exchanged with fresh

media containing 8 µg/ml of polybrene (Catalog Number H9268, Sigma-Aldrich) and lentiviral particles added to an MOI of ~10. After 24 hours, fresh media containing 3 µg/ml of puromycin was added and the cells incubated for 48 hours followed by a second exchange of media containing puromycin, incubation for 48 hours, a third exchange of media and incubation for 24 hours. On day 6, the cells were harvested and half of the cells used to prepare RNA for qRT-PCR and the other half used to prepare protein lysate for western blotting.

RNA Isolation

RNA isolation was performed using a Promega Maxwell 16 Instrument and the Maxwell 16 LEV simplyRNA Cells Kits (Catlog Number AS1270). Briefly, cells were pelleted by centrifugation at 300 x g for 3 minutes and the medium removed. 200 µl of chilled 1-Thioglycerol/Homogenization solution was added to the cell pellet and vortex until dispersed. 200 µl of lysis buffer was added to the cells and the mixture vortex vigorously for 15 seconds. All 400 µl of the lysate was transferred to well 1 of a Maxwell 16 LEV cartridge and 5 µl of DNase I solution added to well 4 of the cartridge. Elution tubes with 40-50 µl of nuclease-free water and LEV plungers were placed in the cartridge and then transferred to the Maxwell 16 Instrument the run started. RNA quality was measured using an Agilent TapeStation Instrument.

qRT-PCR Assay to Monitor mRNA Target Knock-down Efficiency

Reverse transcription was performed using the iScript cDNA Synthsis Kit from BioRAD (Catalog number: 170-8891). Reactions were assembled containing 2µl of 5x iScript reaction mix, 0.5 µl of the iScript reverse transcriptase and 200 ng of the RNA in a total reaction volume of 10 µl. The reactions were incubated for 5 minutes at 25˚C, 30 minutes at 42˚C, 5 minutes at 85˚C and then placed at 4˚C. qPCR assays were assembled using Phusion High-Fidelity DNA Polymerase from NEB (Catalog number: M0530L) and SYBR Green from Invitrogen (Catalog number: S7563) and contained 4µl of 5X Phusion HF Buffer, 0.4 µl of 10 mM dNTPs, 1µl of 10 µM Forward Primer, 1 µl of 10 µM Reverse Primer, 1µl of a 1:20 diulution of the cDNA reaction, 0.2 µl of Phusion DNA Polymerase, and 0.1 µl of 10,000X SYBR Green in a total volume of 20 µl. The reactions were then incubated at 98˚C for 30 seconds followed by 35 cycles of 98˚C for 10 seconds, 58-66˚C (depending on primers used) for 15 seconds, and 72˚C for 10 seconds in a BioRad qPCR machine. Depletion levels were calcluated using the $2^{-\Delta\Delta Ct}$ Method.

Western Blot Assay to Monitor Protein Target Knockdown Efficiency

Cell pellets were resuspended in 100 µl Lysis buffer (50 mM Tris−HCl, pH 7.4; 100 mM NaCl; 1% NP−40; 0.1% SDS; 0.5% sodium deoxycholate) with protease inhibitor (Roche cocktail), vortexed vigorously, and incubated on ice at least 30 min. The lysate was spun at 18,000g for 20 min at 4˚C. The supernatant was collected in a new tube on ice and the protein concentration determined using the Pierce BCA kit. 30−60µg of protein from each sample was diluted into 4X sample buffer and 10X reducing agent (Invitrogen NuPage reagents) and heated at 70˚C for 10 min. The samples were loaded onto 4−12%

Bis−Tris gels for proteins 10−220 kDa, or 8% Tris−glycine gels for proteins over 220 kDa along with 3µl of Licor Odyssey prestained molecular weight marker (928−4000). The gels were run at 200 V for about an hour, until the dye just runs off the gel (less for small molecular weight proteins) in 1X MOPS running buffer with 500 µl of antioxidant to the inner chamber of the buffer tank. The gels were transfered to PVDF for 30 min in transfer buffer, with methanol and antioxidant, using a BioRad Semi−dry transfer apparatus. The membranes were then blocked by incubation in 5−10 mls of Licor Blocking buffer for 1 hour at room temperature. The membranes were then incubated with the RNA binding protein primary antibody (0.2 µg/ml) and the loading control primary antibody (mfg recommended dilution) diluted in Licor block with 0.1% Tween 20 on a rocker at 4˚C overnight. The next day, the membranes were washed 4 times for 5 min each in TBST. The membranes were then incubated with the appropriate secondary antibodies (Rockland Fluorescent TrueBlot anti-rabbit IgG IRDye800 (Catalog number 18−3216−32) for the RBP and Licor IRDye680 secondary antibody for the loading control) that were diluted according to the manufacturers instructions in Licor blocking buffer with 0.1% Tween 20 and 0.01% SDS on a rocker for 30−60 min, at room temperature. The membranes were then washed 4 times for 5 min each in TBST and rinsed once in TBS(noT) and scaned on a Licor Odyssey instrument.

Immunofluorescence characterization

HepG2 cells were seeded in Poly-L-Lysine coated 96-well clear bottom plates (Corning Inc; plate number 3882 half-area microplates), at a concentration of 2,000 cells per well in DMEM + 10%FBS. After 72 hours in standard growth conditions (i.e. 37°C and 5% CO_2), the adhered cells were fixed with 3.7% formaldehyde, permeabilized in PBS + 0.5% Triton X-100 and blocked in PBS + 0.2% Tween-20 (PBST) + 2% BSA, all conducted for 20 minutes at room temperature. Primary antibodies directed against specific RBPs (all rabbit antibodies) and marker proteins (e.g. mouse anti-Fibrillarin, ab4566, 1:200 dilution, Abcam; mouse anti-SC35, GTX11826, 1:200 dilution, GeneTex Inc) were subsequently applied to the cells at a final concentration ranging from 2-10 µg/mL in PBS-BSA (concentration details provided in Table S6), then incubated overnight at 4°C. The cells were next washed 3 times for 10 minutes each with PBST and then incubated with secondary antibodies (Alexa647 donkey anti-rabbit and Alexa488 donkey anti-mouse, both diluted 1:500 in PBST + 2% BSA) for 90 minutes at room temperature. After 3 PBST washes, the cells were counter stained with DAPI for 5 minutes and washed 3 times with PBS and stored in PBS at 4°C. For staining with the Mitotracker dye, the cells were incubated with 100nM of Mitotracker in tissue culture media for 45 minutes at 37°C prior to fixation. Imaging was conducted on an ImageXpress Micro high content screening system (Molecular Devices Inc) using a 40x objective. Exposure times were variable for each RBP antibody and are indicated in Table S6.